AD-A038 955 WISCONSIN UNIV MADISON MATHEMATICS RESEARCH CENTER F/G 12/1
MAXIMUM LIKELIHOOD ESTIMATION OF MULTIVARIATE AUTOREGRESSIVE-MO--ETC(U)
FEB 77 M S PHADKE, G KEDEM DAAG29-75-C-0024
UNCLASSIFIED MRC-TSR-1722 NL

| OF |
AD
A038955

END

DATE
FILMED
5-77

MRC Technical Summary Report # 1722

MAXIMUM LIKELIHOOD ESTIMATION OF MULTIVARIATE AUTOREGRESSIVE-MOVING AVERAGE MODELS

M. S. Phadke and G. Kedem

Mathematics Research Center
University of Wisconsin—Madison
610 Walnut Street
Madison, Wisconsin 53706

February 1977

(Received January 24, 1977)

Approved for public release
Distribution unlimited

Sponsored by

U. S. Army Research Office
P. O. Box 12211
Research Triangle Park
North Carolina 27709

UNIVERSITY OF WISCONSIN - MADISON
MATHEMATICS RESEARCH CENTER

MAXIMUM LIKELIHOOD ESTIMATION OF MULTIVARIATE AUTOREGRESSIVE-MOVING AVERAGE MODELS

M. S. Phadke and G. Kedem

Technical Summary Report # 1722
February 1977

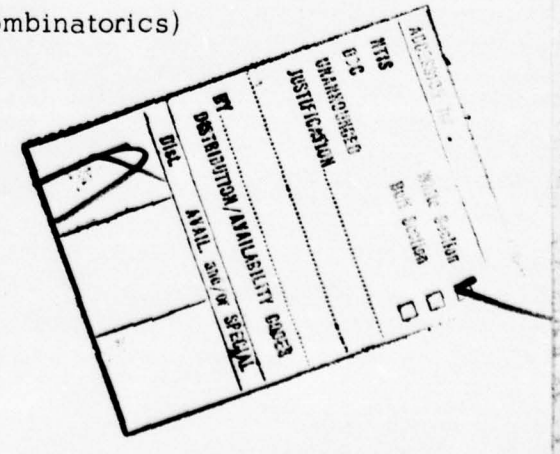
ABSTRACT

Algorithms for computing the exact likelihood function of n successive observation vectors from an s-variate autoregressive moving average process of order (p, q) are developed. A quasi-Newton method is used to maximize the likelihood function with respect to the parameters of the process. Monte Carlo simulations are performed to compare the parameter estimates obtained by maximizing the exact likelihood function versus those obtained by maximizing various approximate forms of the likelihood function.

AMS (MOS) Subject Classification: 62M10

Key Words: likelihood function, multivariate, autoregressive, moving average models

Work Unit Number 4 (Probability, Statistics, and Combinatorics)



Sponsored by the United States Army under Contract No. DAAG29-75-C-0024.

MAXIMUM LIKELIHOOD ESTIMATION OF MULTIVARIATE AUTOREGRESSIVE-MOVING AVERAGE MODELS

M. S. Phadke* and G. Kedem

1. INTRODUCTION

Let $a(t)$ be s-dimensional white noise series with mean zero and diagonal, positive definite covariance matrix Σ. A zero-mean s-dimensional autoregressive-moving average (ARMA) process of order (p,q), denoted by $Y(T)$ is defined by

$$\sum_{j=0}^{p} \phi(j)Y(T-j) = \sum_{j=0}^{q} \theta(j)a(T-j), \tag{1}$$

where $\phi(j)$'s and $\theta(j)$'s are $s \times s$ real matrices; $\phi(0) = I$; and $\theta(0)$ is a lower triangular matrix with diagonal elements equal to unity. Note that there is no loss of generality by the stated assumptions on Σ, $\phi(0)$ and $\theta(0)$.

$$\Phi(z) = \det \left(I + \sum_{j=1}^{p} \phi(j)z^j \right) \tag{2}$$

$$\Theta(z) = \det \left(\theta(0) + \sum_{j=1}^{q} \theta(j)z^j \right) \tag{3}$$

The stationarity condition requires that the zeros of $\Phi(z) = 0$ lie outside the unit circle; and the canonicalness condition requires that the zeros of $\Theta(z) = 0$ lie outside or on the unit circle (see, e.g., Hannan [9]).

*IBM T. J. Watson Research Center, Yorktown Heights, N. Y.

Sponsored by the United States Army under Contract No. DAAG29-75-C-0024.

Let $\gamma(k)$ be the $s \times s$ matrix of covariance at lag k,

$$\gamma(k) = E\,[Y(T)\; Y^T(-k)]$$

Note that $\gamma(k) = \gamma^T(-k)$. Let Y denote the vector formed by n consecutive observations $Y(T)$, $T = 1, \ldots\ldots, n$:

$$Y^T = [Y_1(1)\; Y_2(1) \ldots\ldots Y_s(1) \;\ldots\ldots\; Y_1(n) \ldots\ldots Y_s(n)]$$

The covariance matrix of Y, denoted by Γ, is a symmetric, block Toeplitz matrix whose first row of $s \times s$ blocks is $\gamma(0), \gamma(1), \ldots\ldots, \gamma(n-1)$, i.e.

$$\Gamma = \text{SBT}\;[\gamma(0) \quad \gamma(1) \quad \ldots\ldots \quad \gamma(n-1)]$$

$$= \begin{bmatrix} \gamma(0) & \gamma(1) & \gamma(2) & \ldots\ldots & \gamma(n-1) \\ \gamma(1)^T & \gamma(0) & \gamma(1) & \ldots\ldots & \gamma(n-2) \\ \cdot & & & & \\ \cdot & & & & \\ \cdot & & & & \\ \gamma(n-1)^T & \gamma(n-2)^T & \ldots\ldots & \ldots\ldots & \gamma(0) \end{bmatrix} \tag{4}$$

Assuming that the white noise series $a(T)$ has Gaussian distribution, the log likelihood function (l.l.f.) of the parameters of the model (1) is given by

$$L(\beta) = -\frac{1}{2} \ln |\Gamma| - \frac{1}{2} Y^T \Gamma^{-1} Y \tag{5}$$

where the constant term has been ignored and β stands for all the parameters in $\phi(j)$'s, $\theta(j)$'s and Σ.

The equations obtained by setting the derivatives of the l.l.f. equal to zero are highly nonlinear and are extremely difficult, if not impossible, to solve analytically. Hence, to obtain the maximum likelihood estimates (m.l.e.) of the parameters β, the l.l.f. must be maximized numerically. Further, the computation of the l.l.f. at a given point β is complicated due to the difficulties in computing Γ^{-1} and $\det \Gamma$. Therefore, approximate methods of computing the l.l.f. and maximizing the same have been proposed i) in time domain by Box and Jenkins [4], Astrom [3], Wilson [23], and Phadke [16], and ii) in frequency domain by Hannan [8,9], Anderson [2], Akaike [1] and Parzen [15]. Asymptotically the modifications in l.l.f. have negligible effect. (See, for example, Box and Jenkins [4] and Wahba [22].) However, for finite sample sizes, particularly when the zeros of $\Theta(z) = 0$ are near or on the unit circle, these approximate m.l.e. can have much larger mean squared errors (m.s.e.) than the exact m.l.e.

The objectives of this paper are i) to propose efficient algorithms for computing the exact l.l.f., and ii) to perform simulation for comparing the m.s.e. of exact m.l.e. against the m.s.e. of approximate m.l.e. The problem of computing the exact l.l.f. of univariate ARMA models has been studied by Tiao and Ali [20], Shaman [18,19], Uppuluri and Carpenter [21], Newbold [13] and Ljung [12]. The case of multivariate models was recently investigated by Hillmer [10].

Since the main difficulties of maximum likelihood estimation are due to the moving average part, the case of pure moving average models (p=0) will be presented first. In Section 2 we briefly review the frequency domain and the time domain approximations to the l.l.f. of moving average models. In Section 3 we present three methods for computing the exact l.l.f. of moving average models and compare the relative computational efficiencies of the three methods. Computation of the exact l.l.f. of ARMA(p,q) models is discussed in Section 4. A method for maximizing the l.l.f. is described in Section 5; and Section 6 presents some simulation results for moving average models.

2. APPROXIMATE l.l.f OF MOVING AVERAGE MODELS

For the q^{th} order moving average (MA(q)) model the covariance function is given by

$$\gamma(k) = \gamma^T(-k) = \begin{cases} \sum_{j=0}^{q-k} \theta(j)\, \Sigma\, \theta^T(j+k), & 0 \le k \le q \\ 0, & k > q \end{cases} \tag{6}$$

Hence, the corresponding covariance matrix Γ has block band structure.

$$\Gamma = \text{SBT}\,[\gamma(0)\ \ \gamma(1)\ \ \ldots\ldots\ \ \gamma(q)\ \ 0\ \ 0\ \ \ldots\ldots\ \ 0] \tag{7}$$

$$= \begin{bmatrix} \gamma(0) & \gamma(1) & \ldots\ldots & \gamma(q) & & & \\ \gamma^T(1) & & & & & & \\ \vdots & & & & & & \\ \gamma^T(q) & & & & & & \\ & & & & & & \gamma(q) \\ & & & & & & \vdots \\ & & 0 & & & & \gamma(1) \\ & & & & \gamma^T(q) & \ldots\ldots & \gamma^T(1)\ \ \gamma(0) \end{bmatrix}$$

2.1 Frequency Domain Approximation

Consider the modified covariance matrix Γ_f (ns × ns matrix) defined by

$$\Gamma_f = \text{SBT} \quad [\gamma(0) \;\; \gamma(1) \ldots \;\; \gamma(q) \;\; 0 \ldots 0 \;\; \gamma(q)^T \ldots \gamma(1)^T] \tag{8}$$

$$= \begin{bmatrix} \gamma(0) & \cdots & \gamma(q) & & \gamma^T(q) & \cdots & \gamma^T(1) \\ \vdots & & & & & & \vdots \\ \gamma^T(q) & & & 0 & & & \gamma^T(q) \\ & & & & & & \\ \gamma(q) & & & & & & \gamma(q) \\ \vdots & & 0 & & & & \vdots \\ \gamma(1) & \cdots & \gamma(q) & & \gamma^T(q) & \cdots & \gamma(0) \end{bmatrix}$$

The matrix Γ_f differs from Γ only in the top right corner and the bottom left corner. In addition to being symmetric, block Toeplitz, the matrix Γ_f is also a circulant matrix and hence has the following decompositon:

$$\Gamma_f = X^H F X \tag{9}$$

where X is ns×ns unitary matrix defined by $X = \{X_{jr}\}$ where X_{jr} are s×s matrices,

$$X_{jr} = \frac{1}{\sqrt{n}} I_s \exp(2\pi ijr/n) \tag{10}$$

for j, r = 1, 2,..., n; the superscript H means transpose and complex conjugate; and F is ns×ns block diagonal matrix, $F = \{F_{jr}\}$, where F_{jr} are s×s matrices,

$$F_{jr} = \begin{cases} 0 & j \neq r \\ \sum_{k=-q}^{q} \gamma(k) \exp(-2\pi ikr/n), & j = r \end{cases} \tag{11}$$

for $j, r = 1, 2, \ldots, n$. Note that the diagonal blocks F_{rr} are the same as the spectral densities of the moving average process at equispaced frequencies.

If we approximate the covariance matrix of Y by Γ_f, then the corresponding approximate l.l.f. becomes

$$L_f = -\frac{1}{2}\left\{\sum_{r=1}^{n} \ln[\det F_{rr}] + \sum_{r=1}^{n} w(r)^H F_{rr}^{-1} w(r)\right\} \tag{12}$$

where $w(r)$ is $s \times 1$ vector,

$$w(r) = \sum_{T=1}^{n} Y(T) \exp(2\pi i Tr/n) \tag{13}$$

Note that $w(r)$ are the sample Fourier Transforms of the observed data.

2.2 Time Domain Approximation

Suppose we assume that the innovations $a(0), a(-1), \ldots, a(-q+1)$ are all equal to their expected values, i.e., equal to the zero vector. Then one can recursively compute

$$a(T) = \theta(0)^{-1}\left(Y(T) - \sum_{j=1}^{q} \theta(j)a(T-j)\right) \tag{14}$$

for $T = 1, 2, \ldots, n$. Conditional on $a(0) = a(-1) = \ldots = a(-q+1) = 0$, the l.l.f. of β for given observations Y is given by

$$L_t = -\frac{1}{2}\left\{n \times \ln(\det\Sigma) + \sum_{T=1}^{n} a^T(T)\Sigma^{-1}a(T)\right\} \tag{15}$$

Estimates of β obtained by minimizing L_t will be termed as the approximate time domain m.l.e.

Another way of arriving at the approximate l.l.f. L_t is to approximate the covariance matrix of Y by the following matrix:

$$\Gamma_t = \underline{L}\,\underline{D}\,\underline{L}^T \tag{16}$$

where $\underline{L}$ is ns×ns lower triangular matrix defined by

$$\underline{L} = \begin{bmatrix} \theta(0) & & & & \\ \theta(1) & \theta(0) & & & \\ \vdots & \vdots & \ddots & & \\ \theta(q) & \theta(q-1) & \cdots & \theta(0) & \\ & \theta(q) & \cdots & \theta(1) & \theta(0) \end{bmatrix} \tag{17}$$

for j, r = 1, 2, ..., n; and $\underline{D}$ is (ns×ns) block diagonal matrix defined by

$$\underline{D} = \begin{bmatrix} \Sigma & 0 & 0 & \cdots & \cdots & 0 \\ 0 & \Sigma & 0 & \cdots & \cdots & 0 \\ 0 & 0 & \Sigma & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & & & \vdots \\ 0 & 0 & 0 & \cdots & \cdots & \Sigma \end{bmatrix} \tag{18}$$

3. EXACT l.l.f. OF MOVING AVERAGE MODELS

An efficient algorithm for computing the l.l.f. must exploit to the fullest extent the symmetric, block-band, block-Toeplitz property of the Γ matrix. Further, in order to compute the exact l.l.f. we need not explicity compute Γ^{-1}; but it is enough to be able to compute det Γ and the quadratic form $Y^T\Gamma^{-1}Y$. In the following we propose three techniques for computing the exact l.l.f. Each method utilizes the structure of Γ in a different way: the first method is based on the Cholesky decompostion of the Γ matrix, and the other two methods are based on low rank corrections to the approximate l.l.f. of the frequency domain method and of the time domain method.

3.1 Cholesky Decomposition Method

Since Γ is symmetric, positive definite, (2(q+1)s–1) diagonal band matrix, it has the following Cholesky decomposition (see, e.g., Noble [14], and Conte and deBoor [5]):

$$\Gamma = LDL^T \tag{19}$$

where L is ns×ns lower triangular, (q+1)s diagonal band matrix with the entries of the principal diagonal equal to unity; and D is ns×ns diagonal matrix. The two components of the l.l.f. are given by

$$y^T\Gamma^{-1}y = (L^{-1}y)^T D^{-1} (L^{-1}y), \tag{20}$$

$$\ln (\det \Gamma) = \ln (\det D) = \sum_{r=1}^{n} \ln D_{rr} \tag{21}$$

Thus, the l.l.f. for a given value of the β vector can be computed through the following steps:

(i) Compute the covariances $\gamma(k)$ for $k = 0, 1, ..., q$.

(ii) Compute the factors L and D by a triangularization algorithm for band matrices.

(iii) Compute $L^{-1}y = u$ by a back substitution algorithm for band matrices.

(iv) Compute the weighted inner product $u^TD^{-1}u$ and the ln (det D).

The following considerations are important for reducing the needed storage space and computing time.

1) It is not necessary to store the entire matrix Γ. The desired entries of Γ can be easily obtained from the stored values of $\gamma(0), ..., \gamma(q)$.

2) Steps (ii), (iii) and (iv) need not be carried out sequentially. By interfacing (iii) and (iv) with (ii), one needs only to store current $(q+1)s^2/2$ entries of L and $(q+1)s$ entries of D and u.

3) The fact that Γ has block band structure implies that Γ is a band matrix of variable width. This fact should be utilized to reduce the computation time of steps (ii) and (iii). For small q, the savings thus achieved can be appreciable.

3.2 Frequency Domain Method

Here we use the fact that $\Gamma-\Gamma_f$ is a sparse, low rank (at most 2qs) matrix. The following identity can be easily verified:

$$\Gamma = \Gamma_f + UGCG^TU^T \tag{22}$$

where C, G and U are of dimensions $2qs \times 2qs$, $2qs \times 2qs$ and $ns \times 2qs$ respectively and are defined by:

$$C = \begin{bmatrix} 0 & I_{qs} \\ I_{qs} & 0 \end{bmatrix} \quad ; \quad G = \begin{bmatrix} G^T & 0 \\ 0 & I_{qs} \end{bmatrix}$$

$$U^T = \begin{bmatrix} I_{qs} & 0 & 0 \\ 0 & 0 & I_{qs} \end{bmatrix}$$

where

$$G^T = \begin{bmatrix} \gamma(q)^T & \gamma(q-1)^T & \cdots\cdots & \gamma(1)^T \\ 0 & \gamma(q)^T & \cdots\cdots & \gamma(2)^T \\ . & & & \\ . & & & \\ . & & & \\ 0 & 0 & \cdots\cdots & \gamma(q)^T \end{bmatrix}$$

The matrix C is always nonsingular and Γ_f is nonsingular, except possibly in the case the roots of $\Theta(z) = 0$ have unit modulus. Excluding the cases when Γ_f is singular, by Woodbury's formula (see, e.g., Noble [14] and Householder [11]), we obtain:

$$\Gamma^{-1} = \Gamma_f^{-1} - \Gamma_f^{-1} U G [-C + G U \Gamma_f^{-1} U^T G^T]^{-1} G^T U^T \Gamma_f^{-1} \tag{23}$$

The desired quadratic form in Y is then given by

$$Y^T \Gamma^{-1} Y = Y^T X^H F^{-1} X Y - B^T G [-C + G^T R G]^{-1} G^T B \tag{24}$$

where $B = U^T X^H F^{-1} X Y$ is $2qs \times 1$ real vector and $R = U^T X^H F^{-1} X U$ is $2qs \times 2qs$ real symmetric matrix. Note that the first term on the right hand side of eq. (24) is the same as the approximate quadratic form of the frequency domain method, and we shall call the second term the 'low rank correction term' for obvious reasons. By Woodbury's formula, we also have (see Noble [14]) the following convenient expression for computing det Γ.

$$\det \Gamma = (\det \Gamma_f)(\det C)(\det [-C + G^T R G]) \tag{25}$$

We now propose the following algorithm for computing the l.l.f. for a given value of the β vector: $\frown$

i) Compute the discrete Fourier transform, $w(r)$, $r = 1, 2, \ldots\ldots, n$ of the observed data $Y(T)$ by eq. (13).

ii) Compute the covariances $\gamma(k)$, $k = 1, \ldots\ldots, q$ by Eq. (6).

iii) Compute the spectral density matrices F_{rr}, $r = 1, 2, \ldots\ldots, n$ by eq. (9).

iv) Compute $v(r) = F_{rr}^{-1} w(r)$, $r = 1, 2, \ldots\ldots, n$.

v) Compute

$$b(j) = \begin{cases} \frac{1}{\sqrt{n}} \sum_{r=1}^{n} \exp\left(-\frac{2\pi i r j}{n}\right) v(r), & 1 \le j \le q \\ \frac{1}{\sqrt{n}} \sum_{r=1}^{n} \exp\left(\frac{2\pi i r (n+j-2q)}{n}\right) v(r), & q+1 \le j \le 2q \end{cases} \tag{26}$$

Note that $b(j)$'s are $s \times 1$ real vectors. The vector $\circledast$ is given by

$$B^T = [b^T(1) \quad b^T(2) \ldots\ldots b^T(2q)]$$

vi) Computation of R: Let R be partitioned into $s \times s$ submatrices $R_{j,k}$, $R = \{R_{j,k}\}$, for $j, k = 1, 2, \ldots\ldots$, $2q$. R is a real symmetric matrix and its entries in the upper triangle are given by

$$R_{j,k} = \begin{cases} \frac{1}{n} \sum_{r=1}^{n} \exp\left(\frac{2\pi i r (k-j)}{n}\right) F_{rr}^{-1} & \text{for } j,k = 1,\ldots,q \\ \frac{1}{n} \sum_{r=1}^{n} \exp\left(\frac{2\pi i r (n+k-2q-j)}{n}\right) F_{rr}^{-1} & \text{for } j = 1,\ldots,q;\ k = q+1,\ldots,2q \\ \frac{1}{n} \sum_{r=1}^{n} \exp\left(\frac{2\pi i r (k-j)}{n}\right) F_{rr}^{-1} & \text{for } j,k = q+1,\ldots,2q \end{cases} \tag{27}$$

A careful examination of the above formula reveals that one need only compute $R_{1,k}$ for $k = 1, 2, \ldots\ldots, q$ and $R_{j,q+1}$ for $j = 1, 2, \ldots\ldots, q$. The remaining block entries $R_{j,k}$ are equal to these entries or their transposes.

vii) Compute the approximate quadratic form

$$Q_1 = \sum_{r=1}^{n} w(r)^H v(r) \quad \text{and} \quad \det \Gamma_f.$$

viii) Complete the computation of l.l.f. by appropriate substitutions in eqs. (24), (25) and (5).

Considerable saving in storage space can be achieved by performing steps iii) through viii) in a loop on r and accumulating the various quantities as one goes along. Further, one need only go through half the loop on r since the quantities associated with $r = j$ and $r = n-j$ are complex conjugates. The cases of even and odd n will have to be treated slightly differently.

3.3 Time Domain Method

The low rank correction method in time domain is based on the fact that $\Gamma - \Gamma_t$ is a sparse, low rank (at most qs) matrix. We can easily verify that

$$\Gamma = \Gamma_t + \underline{U}\ \underline{G}\ \underline{C}\ \underline{G}^T\ \underline{U}^T \tag{28}$$

where the matrices $\underline{C}$, $\underline{G}$ and $\underline{U}$ are of dimensions $qs \times qs$, $qs \times qs$, $ns \times qs$, and they are defined by

$$\underline{C} = \begin{bmatrix} \Sigma & 0 & \ldots & \ldots & 0 \\ 0 & \Sigma & \ldots & \ldots & 0 \\ \vdots & & & & \\ 0 & \ldots & \ldots & \ldots & \Sigma \end{bmatrix}$$

$$\underline{G} = \begin{bmatrix} \theta(q) & \theta(q-1) \ldots\ldots & \theta(1) \\ 0 & \theta(q) \ldots\ldots & \theta(2) \\ \vdots & & \\ 0 & 0 \ldots\ldots & \theta(q) \end{bmatrix}$$

$$\underline{U}^T = [I_{qs} \quad 0 \quad 0 \quad \ldots\ldots \quad 0]$$

Since both $\underline{C}$ and Γ_t are nonsingular, by Woodbury's formula, we have:

$$\Gamma^{-1} = \Gamma_t^{-1} - \Gamma_t^{-1} \underline{U}\, \underline{G}\, [\, \underline{C}^{-1} + \underline{G}^T\, \underline{U}^T\, \Gamma_t^{-1}\, \underline{U}\, \underline{G}\,]^{-1}\, \underline{G}^T\, \underline{U}^T\, \Gamma_t^{-1} \tag{29}$$

Thus, the desired quadratic form in Y is given by

$$Y^T \Gamma^{-1} Y = Y^T \underline{L}^{-T} \underline{D}^{-1}\, \underline{L}^{-1} Y - \underline{B}^T \underline{G}\, [\, \underline{C}^{-1} + \underline{G}^T\, \underline{R}\, \underline{G}\,]^{-1}\, \underline{G}^T\, \underline{B} \tag{30}$$

where $\underline{B} = H^T \underline{D}^{-1} \underline{L}^{-1} Y$ is $qs \times 1$ vector, $\underline{R} = H \underline{D}^{-1} H^T$ is $qs \times qs$ symmetric matrix and $H = \underline{L}^{-1}\, \underline{U}$ is $ns \times ns$ matrix. Thus the desired quadratic form is seen to consist of the approximate time domain quadratic form and a low rank correction term. By Woodbury's formula, we also have a convenient formula for computing $\det \Gamma$

$$\det \Gamma = (\det \Gamma_t)\, (\det \underline{C})\, (\det [\underline{C}^{-1}\, \underline{G}^T\, \underline{R}\, \underline{G}]) \tag{31}$$

We now have the following algorithm for computing the l.l.f. for a given value of the β vector.

i) Compute $a(T)$ for $T=1,2,\ldots,n$ by eq. (12) by assuming $a(0) = a(-1) = \ldots = a(-q+1) = 0$

ii) The approximate quadratic form is given by:

$$\sum_{T=1}^{n} a^T(T)\ \Sigma^{-1}\, a(T)$$

iii) Let H be partioned into $s \times s$ submatrices H_{rk}, $H = \{H_{rk}\}$ for $r = 1, 2, \ldots, n$ and $k = 1, 2, \ldots, q$. These submatrices can be computed by the following equations:

$$H_{1,1} = \theta(0)^{-1}$$

$$H_{r,1} = \theta(0)^{-1} \left(\sum_{j=1}^{\min(q,r)} \theta(j)\, H_{r-j,1} \right) \text{ for } r=2,\ldots,n$$

$$H_{r,j} = \begin{cases} 0 & \text{if } r<j \text{ and } j=2,\ldots,q \\ H_{r-j+1,1} & \text{if } r \geq j \text{ and } j=2,\ldots,q \end{cases}$$

iv) Compute

$$\underline{b}(j) = \sum_{r=1}^{n} H^T_{r,j}\, \Sigma^{-1}\, a(r) \qquad \text{for } j=1,\ldots,q.$$

Note that $\underline{b}(j)$'s are $s\times 1$ vectors. The vector $\underline{B}$ is given by:

$$\underline{B} = [\underline{b}(1)^T\ \underline{b}(2)^T \ldots \underline{b}(q)^T]$$

v) Computation of $\underline{R}$: Let $\underline{R}$ be partitioned into $s\times s$ submatrices $\underline{R}_{j,k}$ $\underline{R} = \{\underline{R}_{j,k}\}$ for $j, k = 1, \ldots, q$. R is symmetric matrix and its entries in the upper triangle are given by:

$$\underline{R}_{j,k} = \sum_{r=1}^{n-k-1} H^T_{r+k-j\,,\,1}\, \Sigma^{-1}\, H_{r,1}$$

Note that while computing $\underline{R}_{1,1}$ one get $\underline{R}_{j,j}$, $j = 2, \ldots, q$ as intermediate results for, $j\geq 2$. Similarly, while computing $\underline{R}_{1,2}$ $\underline{R}_{1,3}$, ..., $\underline{R}_{1,q}$ one gets the remaining entiries as intermediate results.

vi) Use Eqs (30), (31) and (5) to complete the computation of the l.l.f.

Similar to the frequency domain method, considerable savings in storage space may be achieved by performing the above steps in a loop on r.

3.4 Comparison of the Three Methods

The three methods of computing the exact l.l.f. utilize the structure of Γ in different ways and hence may be expected to have different computational efficiencies. To facilitate the comparison of the computational effort involved in each method, we give the following formulas for operation counts (OC):

Cholesky Decompostion Method

$$OC = n\,[q^2s^3/2 + qs^2(s+4)/2 + s(s+3)(s+5)/8] \qquad (32)$$

Frequency Domain Method

$$OC = n\,[q(3s)(s+1)+s((8/3)s^2 + (3/2))] \qquad (33)$$

Time Domain Method

$$OC = n[(qs^2(2s+3)+s(s^2+3)/2] \tag{34}$$

where by an operation we mean a multiplication or division of two real numbers. Thus multiplication of two complex numbers is equivalent to four operations. In deriving the above formula terms that do not depend on n have been ignored.

A comparison of the operation counts given by the above equations indicates the following particular choices:

i) $s = 1$: Cholesky method has the smallest OC for $q \leq 4$ and the time domain method has the smallest OC for $q \geq 5$

ii) $s = 2$: Cholesky method is the best for $q \leq 3$ while the frequency domain method is the best for $q \geq 4$

iii) $s = 3$, $s = 4$: Cholesky method is the best for $q \leq 2$ and the frequency domain method has the smallest OC for $q \geq 3$.

4. ESTIMATION OF ARMA (p,q) MODELS

Consider the ARMA (p,q) process as being defined by Eq.(1) and Z(T) as being defined by:

$$Z(T) = \sum_{j=0}^{p} \phi(j)Y(T-j) \tag{35}$$

Then Z(T) is a MA(q) process. We have

$$p\,(Y(p+1), Y(p+2), \ldots, Y(n); Y(1), \ldots, Y(p), \beta)$$

$$= p\,(Z(p+1), \ldots, Z(n); Y(1), \ldots, Y(p), \beta) \tag{36}$$

since the Jacobian of the linear transformation from Z(T)'s to Y(T)'s is equal to unity. Thus, the l.l.f. of Y(p+1), ..., Y(n) with the first observations Y(1), ..., Y(p) considered fixed, can be computed by one of the three methods described in the previous section. For obvious reasons, the efficiency of the parameter estimates obtained by maximizing the above conditional likelihood is at least (n–p)/n. Thus, for most practical situation where $n >> p$, these conditional estimates may prove to be satisfactory.

The exact likelihood function may be computed using the Bayes's theorem:

$$p(Y;\beta) = p(Y(1),\ldots, Y(p);\beta) \times p(Y(p+1),\ldots, Y(n); Y(1),\ldots, Y(p), \beta) \quad (37)$$

Computation of $p(Y(1), \ldots, Y(p); \beta)$ involves computing the covariance matrices $\gamma(0), \ldots, \gamma(p-1)$ of the ARMA(p,q) process which is quite complicated. So for most practical situations where $n >> p$, one may elect to maximize the conditional likelihood function of Eq. (36) instead of the exact likelihood function given by Eq. (37).

5. OPTIMIZATION METHOD

A quasi-Newton method due to Gill and Murray [7] for unconstrained optimization of general nonlinear functions may be used to maximize the l.l.f. The algorithm computes the needed derivatives of the objective function internally by finite difference approximation and updates the Hessian by Daviden-Fletcher-Powell correction formula [6]. The convergence rate of the quasi-Newton method is superlinear.

Two Stage Optimization: Since computation of the exact l.l.f is much more expensive than computation of either of the two approximations to the l.l.f., it is advisable to first maximize an approximate l.l.f. and then use the resulting estimates and the Hessian as starting values for maximizing the exact l.l.f.

Maximization of the approximate l.l.f.: For maximization of the approximate l.l.f. one may take the starting values to be 'no prior information' values viz. $\phi(1) = \phi(2) = \ldots = \phi(p) = 0$, $\theta(0) = I$, $\theta(j) = 0$ for $j = r, \ldots, q$, and the diagonal entries Σ_{jj} of Σ equal to the variances of the respective $Y_j(T)$ series. Or one may use the procedures described by Box and Jenkins [4] for univariable time series and by Hannan [8,9] for multivariate cases. Our maximization of the approximate l.l.f. by the quasi-Newton method with 'no prior information' starting values has been very encouraging.

Positive definiteness of β: While maximizing the approximate or the exact l.l.f., one must ensure that β remains positive definite. By our formulation of the ARMA models with $\theta(0)$ as lower triangular matrix and β as diagonal, the check for positive definiteness becomes very simple. Further in the numerous simulations we performed, it was observed that requiring the diagonal entries of Σ to be positive was redundant, since the objective function attaches as natural penalty against excusions into nonpositive values of Σ.

Canonicalness of the optimum point: Performing constrained optimization to obtain canonical values of θ parameters is an impractical task. Thus, one may perform an unconstrained optimization, check for canonicalness and if necessary perform the transformations developed by Rozanov [17] and illustrated by Phadke [16] for obtaining equivalent canonical model.

6. SIMULATION RESULTS

Monte Carlo simulations were performed to compute the mean squared errors (m.s.e.) of parameter estimates obtained by maximizing the two approximate forms of the l.l.L. and those obtained by maximizing the exact l.l.f. Table 1 gives the m.s.e. for the parameters of the univariate MA(1) model obtained by simulating 100 time series of 100 observations each with $\theta(1) = .95$ and $\sigma_a^2 = 1$. Table 2 gives similar simulation results for the univariate MA(2) model with $\theta(1) = 0$, $\theta(2) = .95$ and $\sigma_a^2 = 1$; and Table 3 for the bivariate MA(1) model with

$$\theta(0) = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}, \theta(1) = \begin{bmatrix} .98 & 1.96 \\ 0 & .98 \end{bmatrix} \quad \Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

From Tables 1, 2 and 3 we observe that the time domain approximate method gives considerably smaller m.s.e. than the corresponding m.s.e. for the frequency domain approximate method, while the exact likelihood method gives much smaller m.s.e. than either of the two approximate methods.

Simulations were also performed for the cases when all the roots of $\det \Theta(z) = 0$ lie away from the unit circle. As expected, it was found that the two approximate methods and the exact method gave similar m.s.e.

7. CONCLUSIONS

i) Three algorithms are proposed for computing the exact l.l.f. of the multivariate moving average processes. Formulas for operation counts of the algorithms are given as a guide in selecting the best application for a given problem. It is then shown how any one of these algoirithms can be modified for computing the exact l.l.f. of multivariate autoregressive moving average processes.

ii) Monte Carlo simulations are performed to compare the m.s.e. of estimates from the time domain and frequency domain approximate methods and the exact likelihood method. The simulations show that i) the exact likelihood method has the smallest m.s.e. and ii) the time domain approximation gives smaller m.s.e. than the frequency domain approximation.

8. ACKNOWLEDGMENTS

A major portion of the research was carried out when Dr. Phadke was a Research Associate in the Statistics Department and the Mathematics Research Center at the University of Wisconsin–Madison. The research was supported in part by the U.S. Army under Contract No. DAAG29-75-C-0024.

Table 1 Mean Squared Errors of Estimates for the Univariate MA(1) Model: Results of 100 simulations with $\theta(1) = .95$, $\sigma_a^2 = 1$ and $n = 100$.

	Time Domain Approx.	Freq. Domain Approx.	Exact
$\theta(1)$	.00695	.01472	.00180
σ_a^2	.00671	.01481	.00107

Table 2. Mean Squared Errors of Estimates for the Univariate MA(2) Model: Results of 100 simulations with $\theta(1) = 0$, $\theta(2) = .95$, $\sigma_a^2 = 1$ and $n = 100$.

	Time Domain Approx.	Freq. Domain Approx.	Exact
$\theta(1)$	.00360	.00468	.00241
$\theta(2)$	.00916	.02442	.00221
σ_a^2	.01012	.02641	.00191

Table 3. Mean Squared Errors of Estimates for the Bivariate MA(1) Model: Results of 100 simulations with

$$\theta(0) = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}, \theta(1) = \begin{bmatrix} .98 & 1.96 \\ 0 & .98 \end{bmatrix}, \Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

	Time Domain Approx.	Freq. Domain Approx.	Exact
$\theta_{21}(0)$	.01544	.01799	.00970
$\theta_{11}(1)$	.05580	.06883	.03940
$\theta_{21}(1)$	.00917	.01034	.01029
$\theta_{12}(1)$	.00893	.01408	.00145
$\theta_{22}(1)$	.00696	.01106	.00066
σ_1^2	.05298	.08557	.00251
σ_2^2	.00801	.01404	.00266

REFERENCES

1. Akaike, Hirotugu (1973), "Maximum likelihood identification of Gaussian autoregressive moving average models", *Biometrika*, **60**, 255-265.

2. Anderson, T. W. (1975), "Maximum likelihood estimation of parameters of an autoregressive process with moving average residuals and other convariance matrices with linear structure", *Annals of Statistics*, 3.

3. Astrom, K. J. (1970), *Introduction to Stochastic Control Theory*, Academic Press, New York and London.

4. Box, George E. P., and Jenkins, Gwilym M. (1970), *Time Series Analysis Forecasting and Control*, Holden-Day, Inc., San Francisco.

5. Conte, S. P. and deBoor, C. (1972), *Elementary Numerical Analysis*, McGraw Hill Inc.

6. Fletcher, R. and Powell, M. J. D. (1963), "A rapidly convergent descent method for minimization", *Comput. J.*, **6**, 163-168.

7. Gill, P. E. and Murray, W. (1972), "Quasi-Newton methods for unconstrained optimization", *J. of the Inst. of Math. and its Appl.*, **9**, 91-108.

8. Hannan, E. J. (1969), "The estimation of mixed moving average autoregressive systems", *Biometrika*, **56**, 579-593.

9. Hannan, E. J. (1970), *Multiple Time Series*, John Wiley & Sons, Inc., New York.

10. Hillmer, S. (1976), Ph.D. Thesis, Statistics Department, University of Wisconsin-Madison.

11. Householder, A. S. (1964), *The Theory of Matrices in Numerical Analysis*, Dover, New York.

12. Ljung, G. (1976), Ph.D. Thesis, Statistics Department, University of Wisconsin-Madison.

13. Newbold, P. (1974), "The exact likelihood function for a mixed autoregressive moving average process", *Biometrika*, **61**, 423-426.

14. Noble, B. (1969), *Applied Linear Algebra*, Prentice Hall, New Jersey.

15. Parzen, Emanuel (1971), "Efficient estimation of stationary time series mixed schemes", *Bulletin of the International Statist. Inst.*, **44**, 315-319.

16. Phadke, M. S. (1973), "Multiple time series modeling and system identification with applications", Ph.D. thesis, Mechanical Engineering Department, University of Wisconsin-Madison.

17. Rozanov, Y. A. (1967), *Stationary Random Processes*, Holden Day, San Francisco.

18. Shaman, P. (1969), "On the inverse of the covariance matrix of a first order moving average", *Biometrika*, **56**, 595-600.

19. Shaman, P. (1973), "On the inverse of the covariance matrix for an autoregressive moving average process", *Biometrika*, **60**, 193-196.

20. Tiao, G. C. and Ali, M. M. (1971), "Analysis of correlated random effects: linear model with two random components", *Biometrika*, **58**, 37-51.

21. Uppuluri, V. R. R. and Carpenter, J. A. (1969), "The inverse of a matrix occurring in the first order moving average models", *Sankhya*, ser A, **31**, 79-82.

22. Wahba, Grace (1968), "On the distribution of some statistics useful in the analysis of jointly stationary time series", *Annals of Mathematical Statistics*, **39**, 1849-1862.

23. Wilson, G. T. (1973), "The estimation of parameters in multivariate time series models". *J.R.S.S., Ser. B*, 76-85.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: 1722

2. GOVT ACCESSION NO.

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle): MAXIMUM LIKELIHOOD ESTIMATION OF MULTIVARIATE AUTOREGRESSIVE-MOVING AVERAGE MODELS

5. TYPE OF REPORT & PERIOD COVERED: Summary Report, - no specific reporting period

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s): M. S. Phadke and G. Kedem

8. CONTRACT OR GRANT NUMBER(s): DAAG29-75-C-0024

9. PERFORMING ORGANIZATION NAME AND ADDRESS: Mathematics Research Center, University of Wisconsin, 610 Walnut Street, Madison, Wisconsin 53706

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 4 (Probability, Statistics, and Combinatorics)

11. CONTROLLING OFFICE NAME AND ADDRESS: U. S. Army Research Office, P.O. Box 12211, Research Triangle Park, North Carolina 27709

12. REPORT DATE: February 1977

13. NUMBER OF PAGES: 19

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): 22p.

15. SECURITY CLASS. (of this report): UNCLASSIFIED

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

MRC-TSR-1722

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

likelihood function
multivariate
autoregressive
moving average models

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Algorithms for computing the exact likelihood function of n successive observation vectors from an s-variate autoregressive moving average process of order (p, q) are developed. A quasi-Newton method is used to maximize the likelihood function with respect to the parameters of the process. Monte Carlo simulations are performed to compare the parameter estimates obtained by maximizing the exact likelihood function versus those obtained by maximizing various approximate forms of the likelihood function.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)